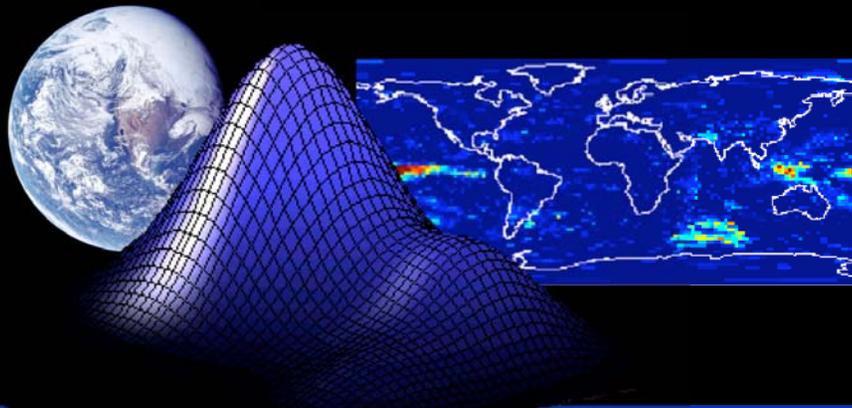
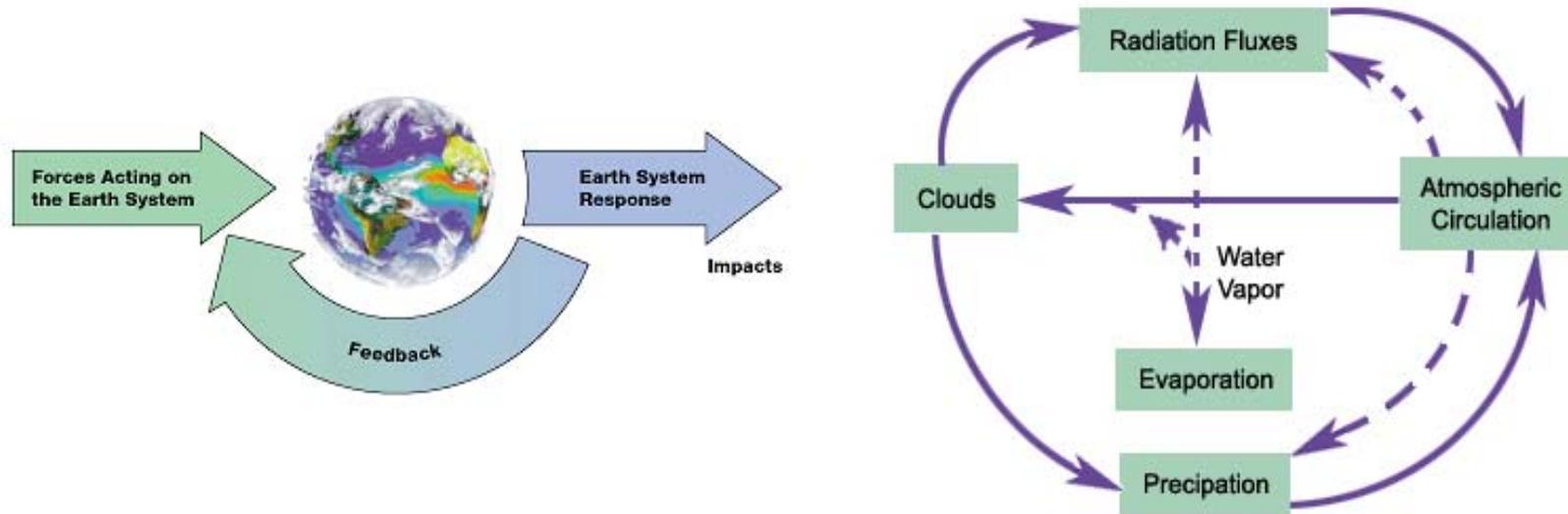

Information-Theoretic Methods for Identifying Relationships among Climate Variables

Kevin H. Knuth
Departments of Physics and Informatics
University at Albany (SUNY)



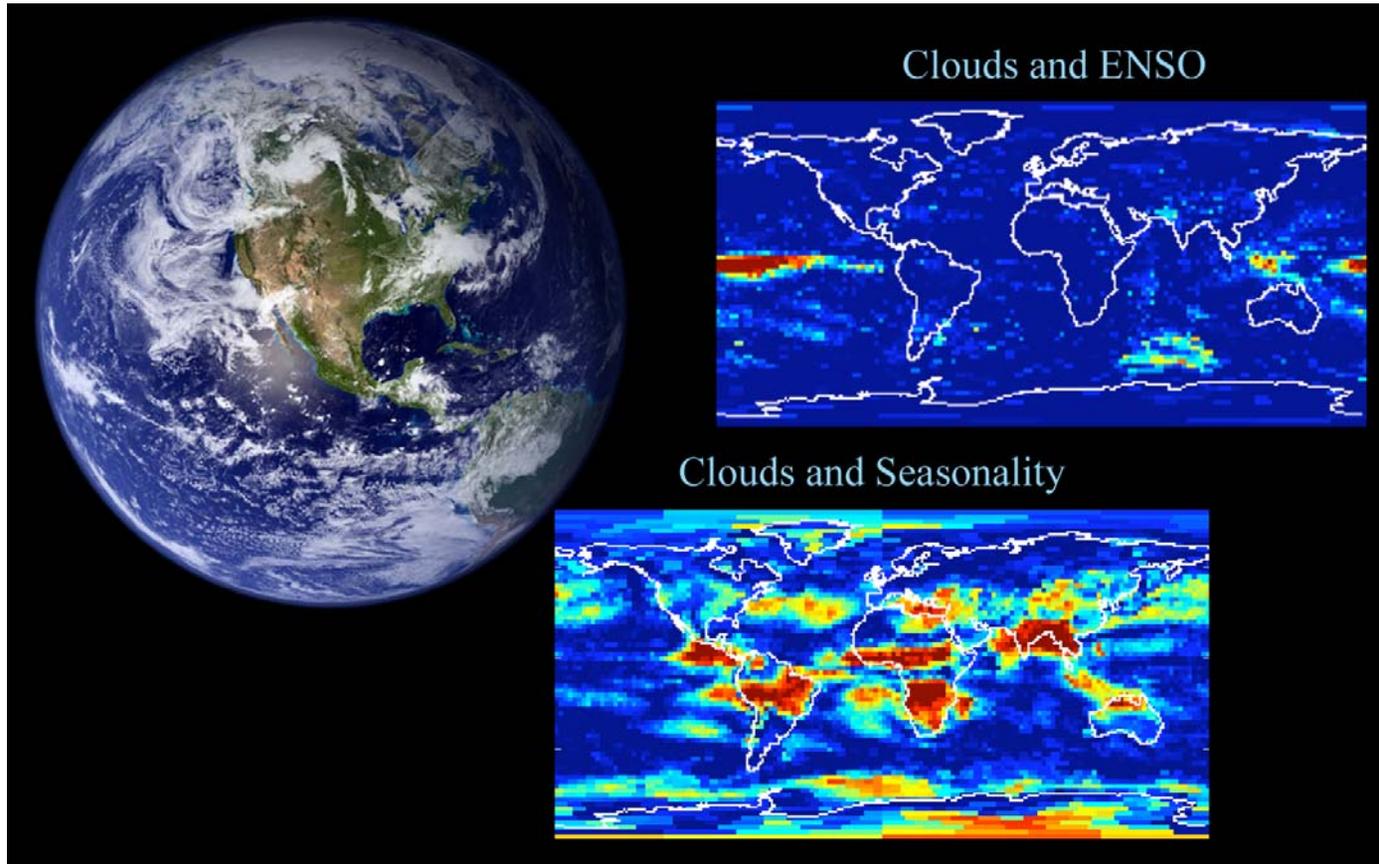
Forcings and Feedback



NASA's goal is to understand the observed Earth climate variability, and determine and predict the climate's response to both natural and human-induced forcing.

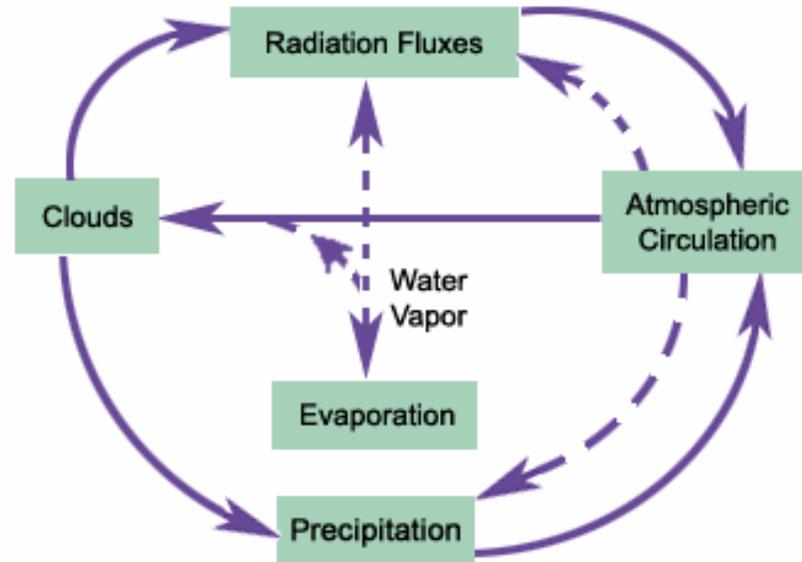
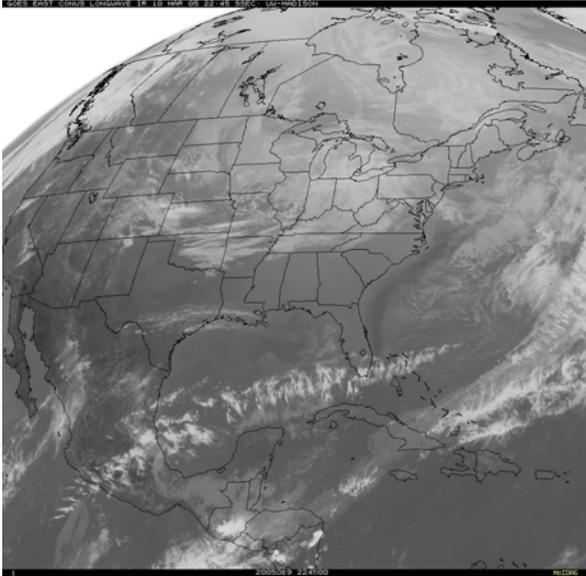
The basic idea is that changes in one climate subsystem will cause or force responses in other subsystems. These responses in turn feed back to force other subsystems.

Relevant Variables



One of the greatest difficulties in this field is the identification of the **RELEVANT VARIABLES**.

Causal Interactions



Once relevant variables are identified, one can begin examining their **CAUSAL INTERACTIONS**, which implement forcings and feedbacks.

However, these calculations are useless without **error bars** that indicate our degree of **uncertainty**. Much of the work we propose to do is aimed toward quantifying our uncertainties.

Outline

Describing Systems

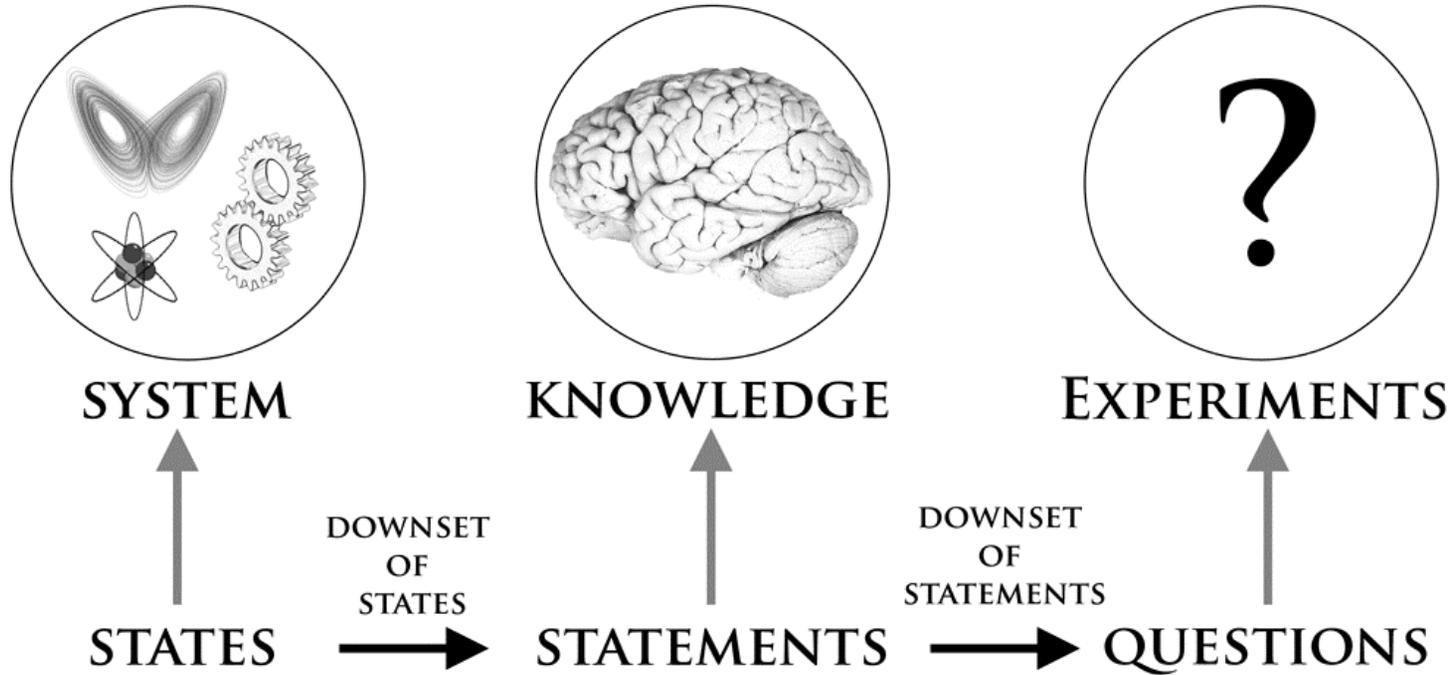
Information Theory

Histogram Methods

Mixture of Gaussians

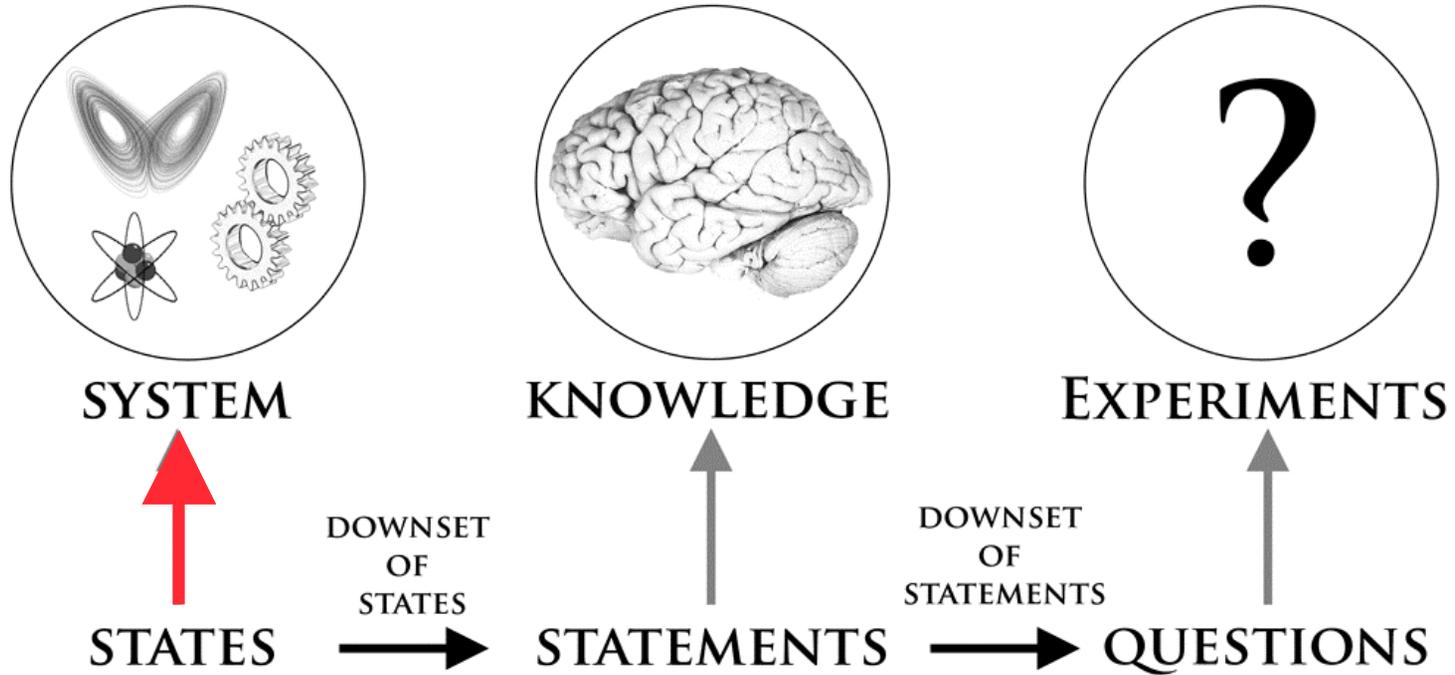
Deployment Efforts (Ongoing)

Describing Physical Systems



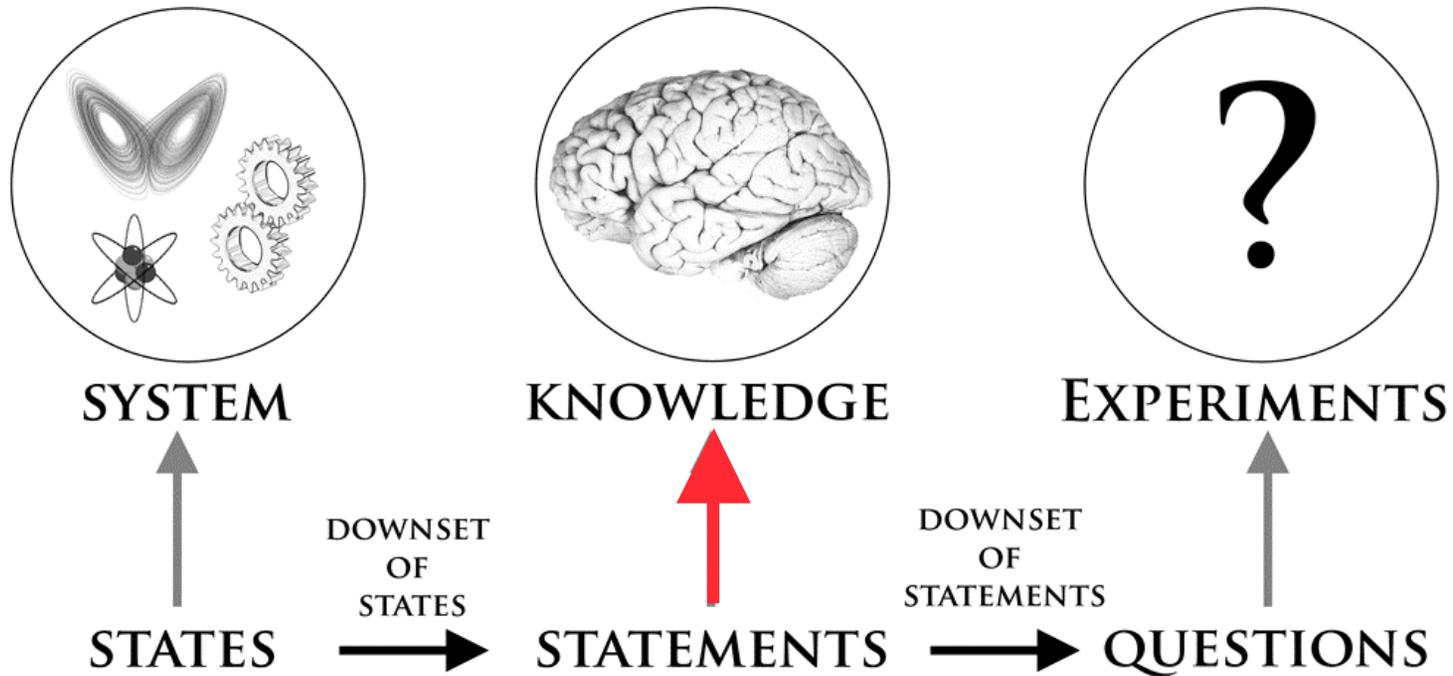
Knuth, *MaxEnt* 2008

Describing Physical Systems



States are the most basic description of a system.

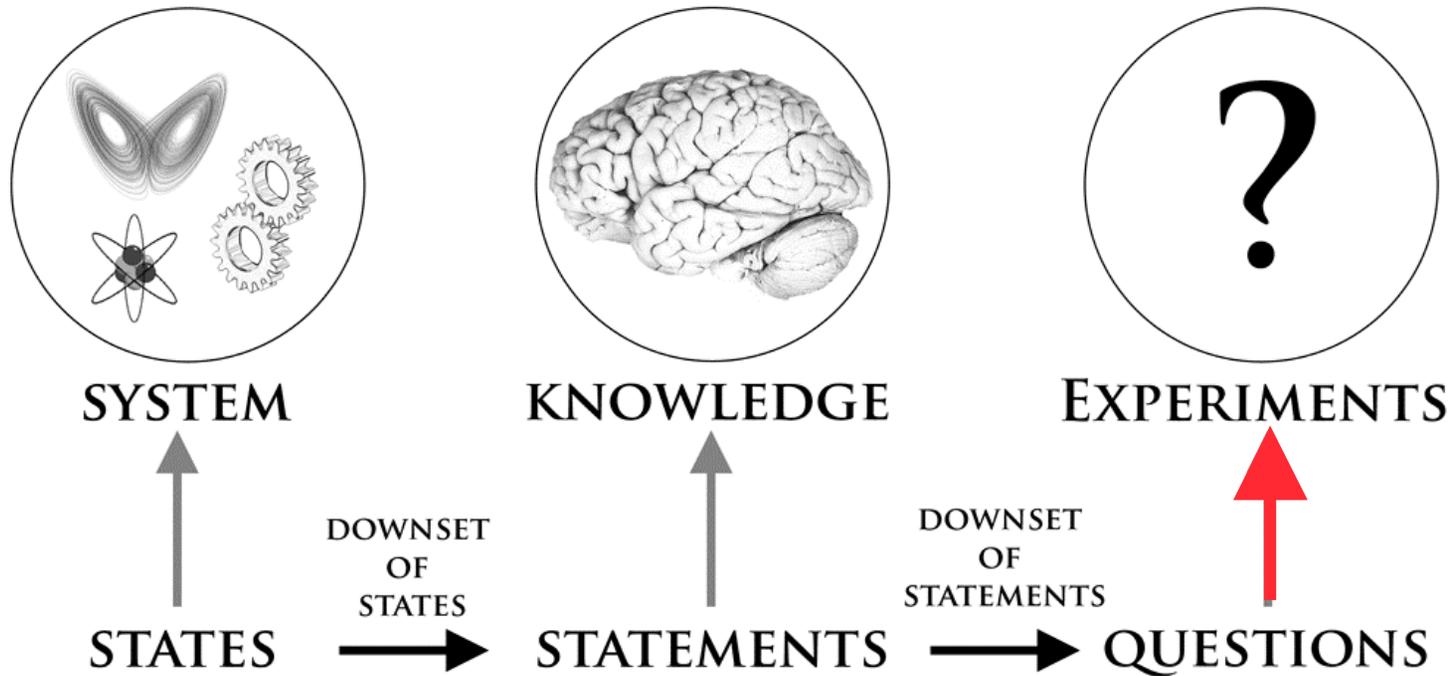
Describing Physical Systems



Statements are sets of potential states.

The space of all statements is the **hypothesis space**

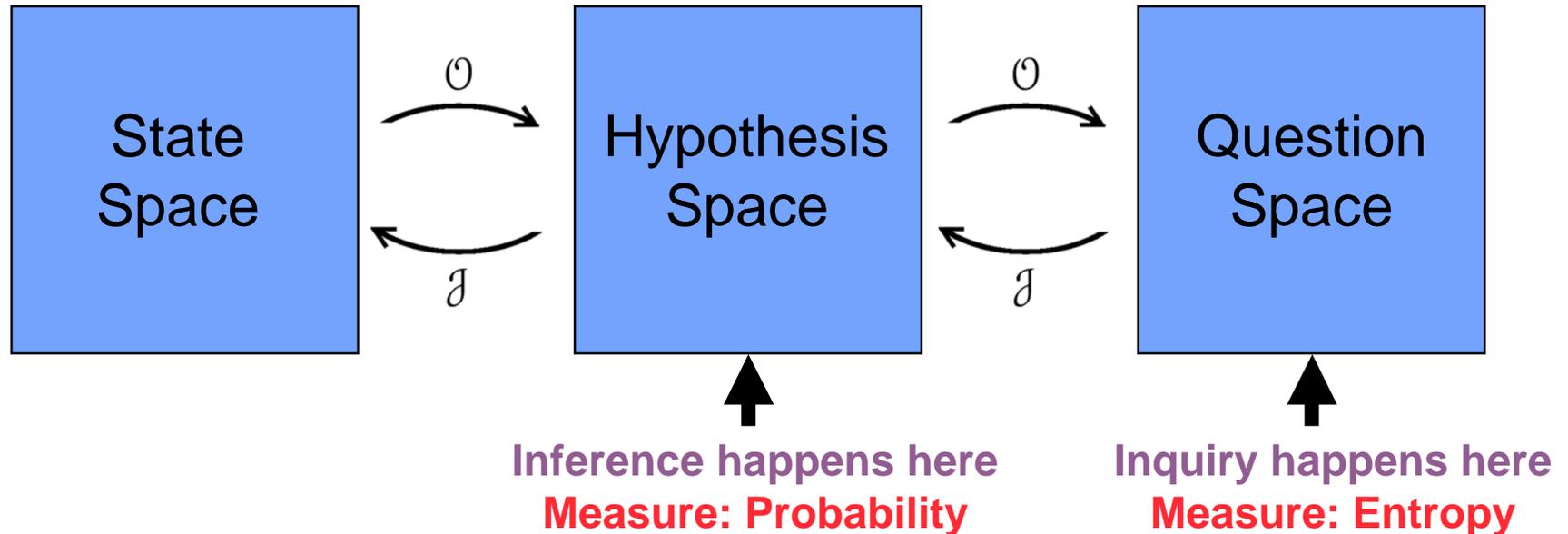
Describing Physical Systems



Questions are sets of potential statements.

When a question has been reduced to a single statement, it has been answered.

Three Spaces



These three spaces enable us to:

- Describe the system
- Describe what we know about the system
- Describe what we could know about the system

Outline

Describing Systems

Information Theory

Histogram Methods

Mixture of Gaussians

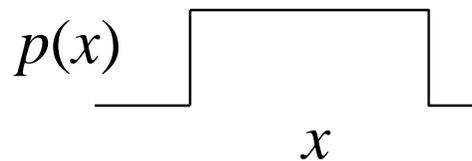
Deployment Efforts (Ongoing)

Expectation and Surprise

When Henry was born, he had no information about the world.

All things were essentially equally probable.

He was equally surprised by everything.



All states equally probable.

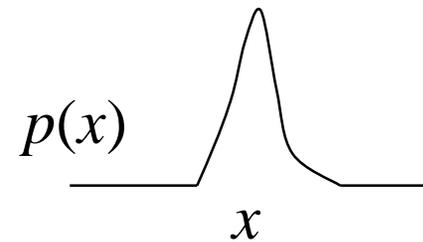


Expectation and Surprise



Henry now has some idea that some events are more probable than others.

He is now sometimes surprised.



Some states rarely occur!

Entropy

We use x to denote the state of the system out of a set of possible states X

The **surprise** is large for improbable states and small for probable states.

$$h(x) = \log \frac{1}{p(x)}$$

Averaging this quantity over all of the possible states of the system gives a measure of our knowledge about the state of the system

$$H(X) = \sum_{x \in X} p(x) \log \frac{1}{p(x)} = - \sum_{x \in X} p(x) \log p(x)$$

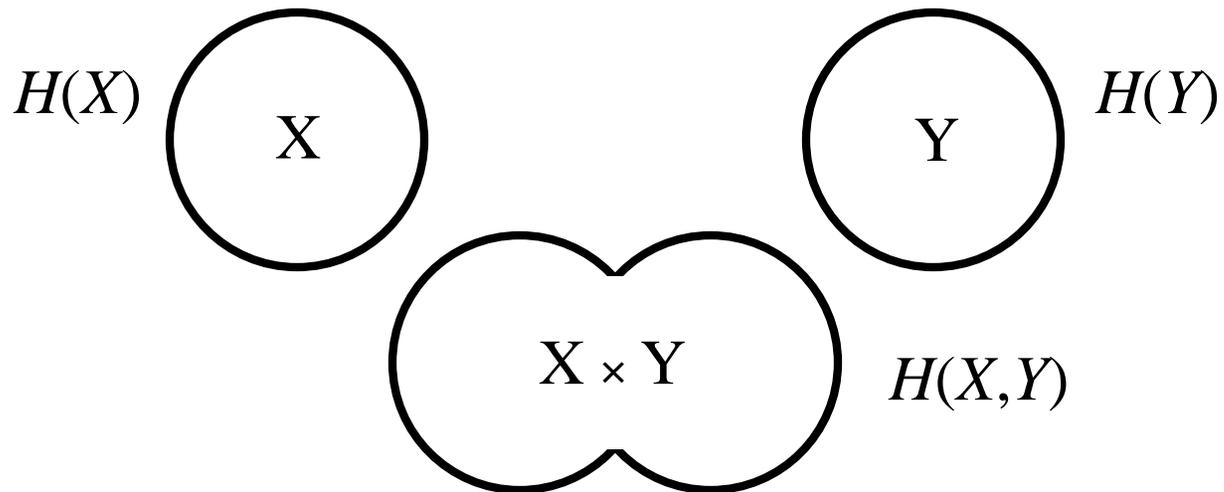
which is called the **entropy**.

Joint Entropy

If the system states can be described with multiple parameters, the entropy is computed by averaging over all possible states

$$H(X,Y) = - \sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x,y)$$

This is called the **Joint Entropy**, since it describes the entropy of the states of X and Y , which jointly describe the system. You can think of X and Y as representing subsystems of the original system $X \times Y$.

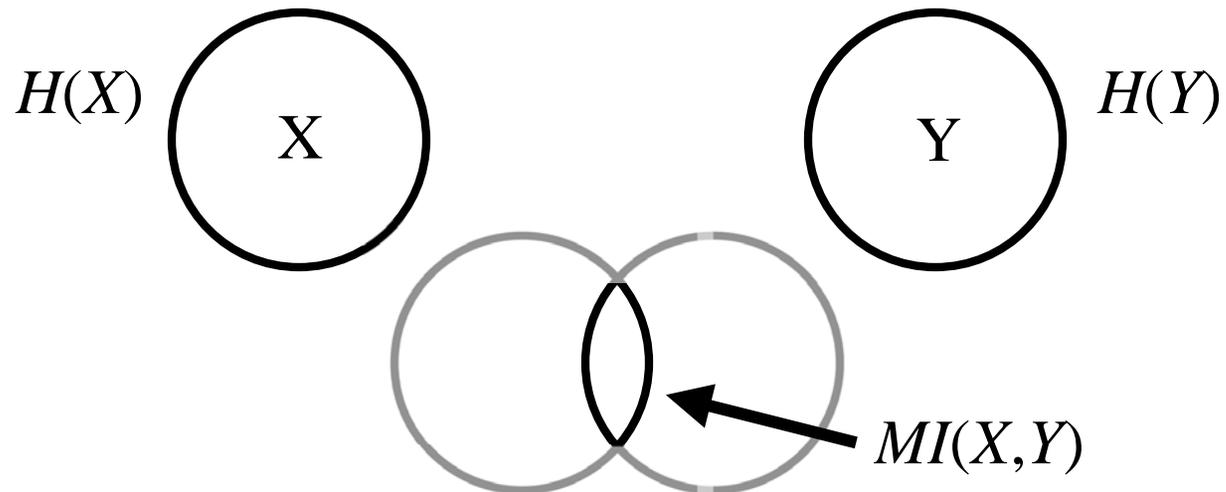


Mutual Information

In this case, an important quantity is the difference of entropies,

$$MI(X, Y) = H(X) + H(Y) - H(X, Y)$$

This is called the **Mutual Information** (MI) since it describes the amount of information that is shared between the two subsystems.



Identifying Relationships

If two climate variables are independent, then the joint entropy is

$$H(X,Y) = H(X) + H(Y)$$

which gives a mutual information of zero, since

$$\begin{aligned} MI(X,Y) &= H(X) + H(Y) - H(X,Y) \\ &= H(X) + H(Y) - [H(X) + H(Y)] \\ &= 0 \end{aligned}$$

While mutual information can identify dependencies, it cannot determine the causal nature the interaction.

Relation to Probability Densities

The mutual information can also be written as

$$MI(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

Which highlights the fact that this is about the **probability density** of the states of the system.

Note that if the joint probability density $p(x, y)$ can be factored into $p(x)p(y)$, then the mutual information is zero and the two systems X and Y are independent.

Transfer Entropy

Recently, Schreiber (2000) introduced a novel information-theoretic quantity called the **Transfer Entropy** (TE). Consider two subsystems X and Y , with data in the form of a two time series of measurements

$$X = \{x_1, x_2, \dots, x_t, x_{t+1}, \dots, x_n\}$$

$$Y = \{y_1, y_2, \dots, y_s, y_{s+1}, \dots, y_n\}$$

then the transfer entropy can be written as

$$T(X_{t+1} | X_t, Y_s) = -H(X_t) + H(X_t, Y_s) + H(X_t, X_{t+1}) - H(X_t, X_{t+1}, Y_s)$$

which describes the degree to which information about Y allows one to predict future values of X . This is then a measure of the causal influence that the subsystem Y has on the subsystem X .

Estimating Information-Theoretic Quantities

Develop proven tools that will allow researchers to identify relevant variables, and to quantify and characterize their causal interactions.

The basic procedure is straightforward:

Estimating Information-Theoretic Quantities

Develop proven tools that will allow researchers to identify relevant variables, and to quantify and characterize their causal interactions.

The basic procedure is straightforward:

1. Estimate the probability density from which the data were sampled.

Estimating Information-Theoretic Quantities

Develop proven tools that will allow researchers to identify relevant variables, and to quantify and characterize their causal interactions.

The basic procedure is straightforward:

1. Estimate the probability density from which the data were sampled.
2. Using this probability density, estimate the various necessary entropies.

Estimating Information-Theoretic Quantities

Develop proven tools that will allow researchers to identify relevant variables, and to quantify and characterize their causal interactions.

The basic procedure is straightforward:

1. Estimate the probability density from which the data were sampled.
2. Using this probability density, estimate the various necessary entropies.

Extremely difficult since not only are we interested in the **values** of these quantities, but we are also interested in the **associated uncertainties** of our estimates.

Outline

Describing Systems

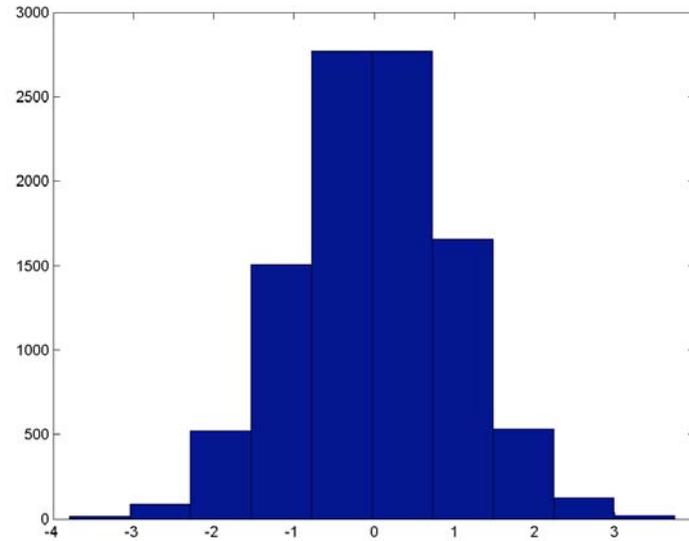
Information Theory

Histogram Methods

Mixture of Gaussians

Deployment Efforts (Ongoing)

Histograms as Probability Density Models

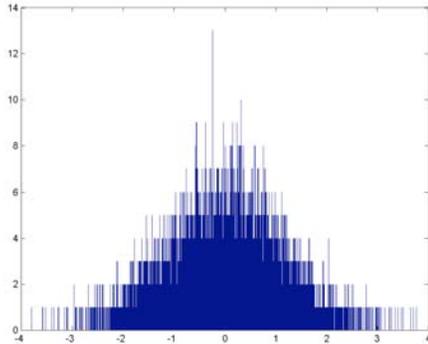


Histograms can be viewed as simple models of the probability density from which the data were sampled.

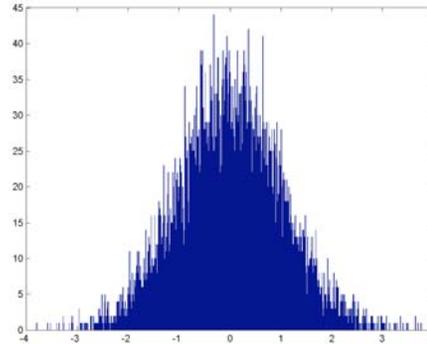
They are convenient since they have regions of constant probability.

Histograms

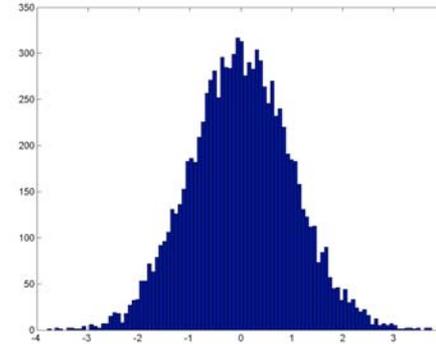
$N = 10000, M = 10000$



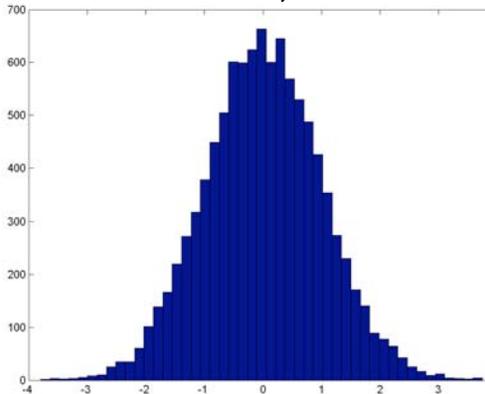
$N = 10000, M = 1000$



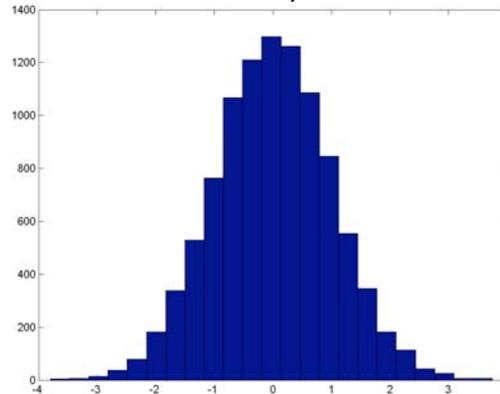
$N = 10000, M = 100$



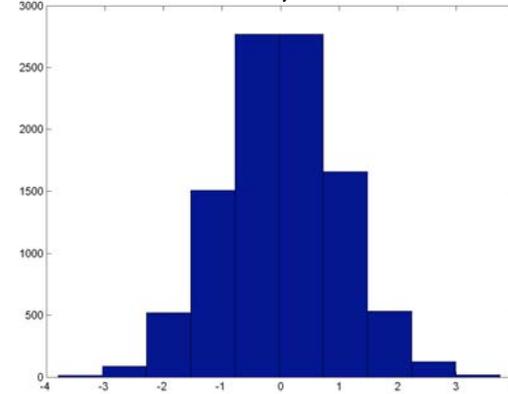
$N = 10000, M = 47$



$N = 10000, M = 23$



$N = 10000, M = 10$



The histogram should contain only details warranted by the data.
But how do we choose the Number of Bins?

Posterior for the Number of Bins

By integrating over all possible bin probabilities, we can derive the posterior probability of the number of bins given the data.

$$p(M | \mathbf{d}, I) \propto \left(\frac{M}{V}\right)^N \frac{\Gamma\left(\frac{M}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^M} \frac{\prod_{k=1}^M \Gamma\left(n_k + \frac{1}{2}\right)}{\Gamma\left(n_1 + b_1 + \frac{3}{2}\right)}$$

It is easier to **find the number of bins that maximizes the logarithm of the posterior probability**

$$\log p(M | \mathbf{d}, I) =$$

$$N \log M + \log \Gamma\left(\frac{M}{2}\right) - M \log \Gamma\left(\frac{1}{2}\right) - \log \Gamma\left(N + \frac{M}{2}\right) + \sum_{k=1}^M \log \Gamma\left(n_k + \frac{1}{2}\right) + K$$

where K is the implicit proportionality constant.

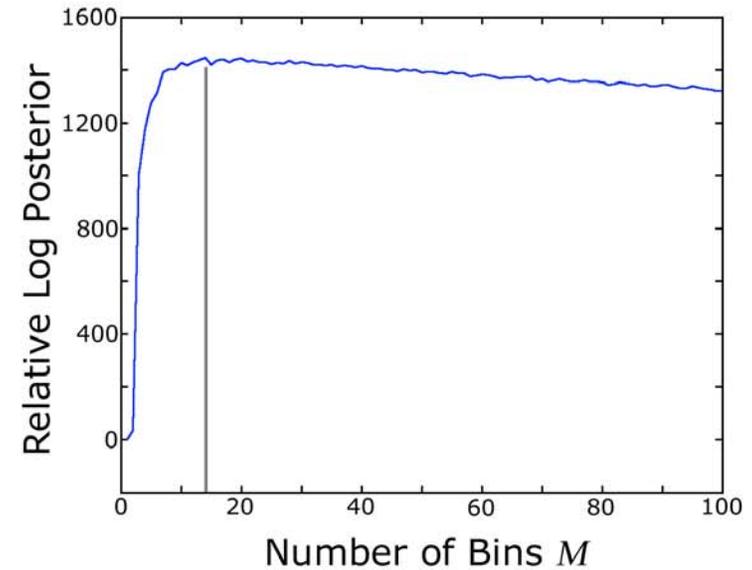
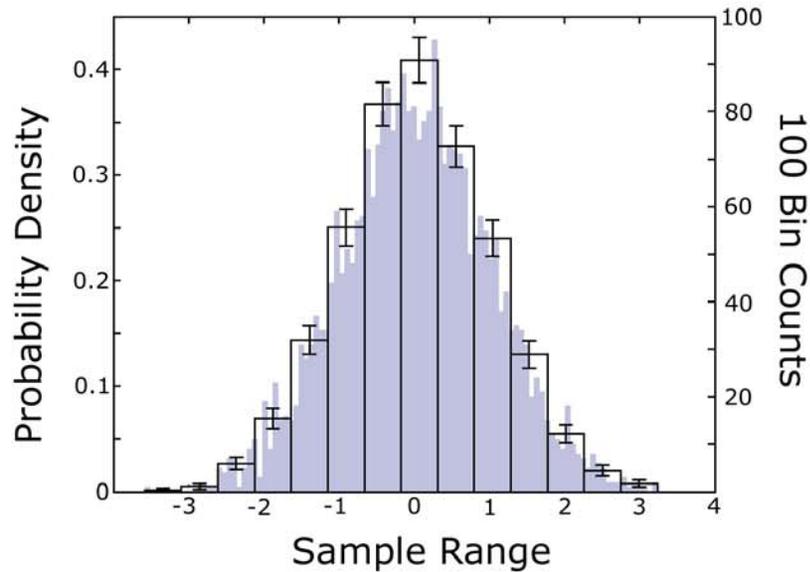
optBINS Algorithm

```
function optM = optBINS(data,minM,maxM)
if size(data)>2 | size(data,1)>1
    error('data dimensions must be (1,N)');
end
N = size(data,2);

% Loop through the different numbers of bins
% and compute the posterior probability for each.
logp = zeros(1,maxM);
for M = minM:maxM
    n = hist(data,M); % Bin the data (equal width bins here)
    p = 0;
    for k = 1:M
        p = p + gammaln(n(k)+0.5);
    end
    logp(M) = N*log(M) + gammaln(M/2) - M*gammaln(1/2) - gammaln(N+M/2) + p;
end
[maximum, optM] = max(logp);
return
```

Knuth, 2008, In review at *Technometrics*

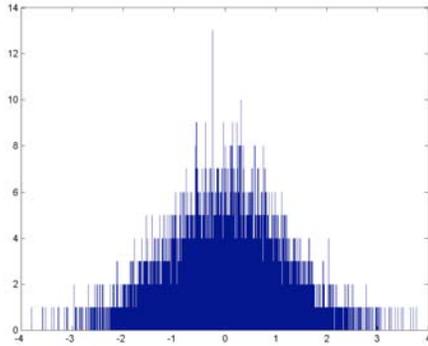
Optimal Histograms



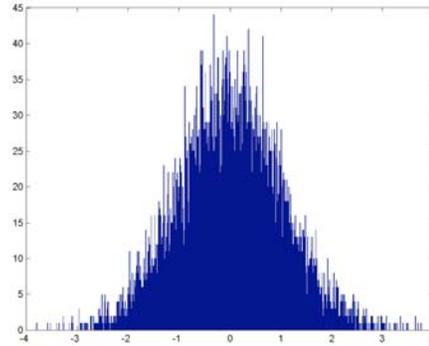
Optimal Binning for $N = 3000$ Gaussian distributed data points: **$M = 14$**

The Optimal Histogram

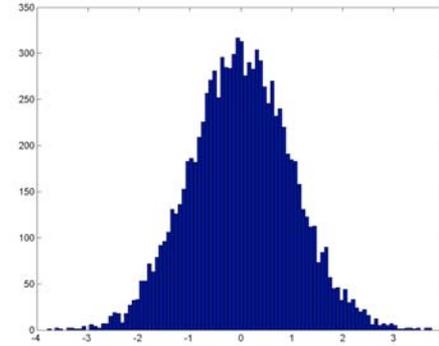
$N = 10000, M = 10000$



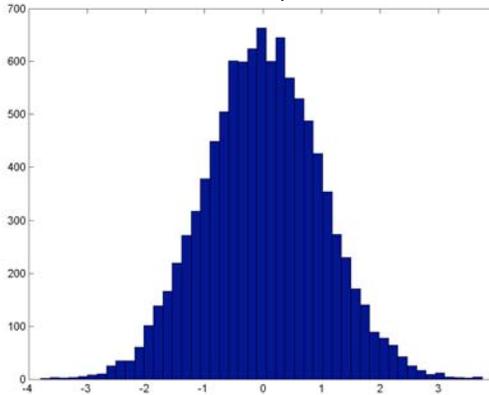
$N = 10000, M = 1000$



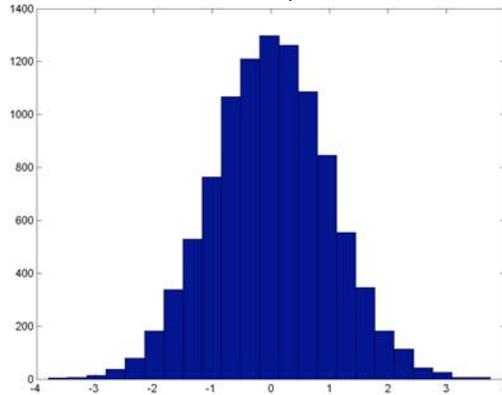
$N = 10000, M = 100$



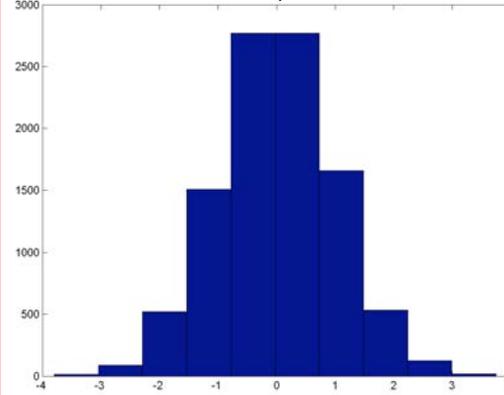
$N = 10000, M = 47$



$N = 10000, M = 23$



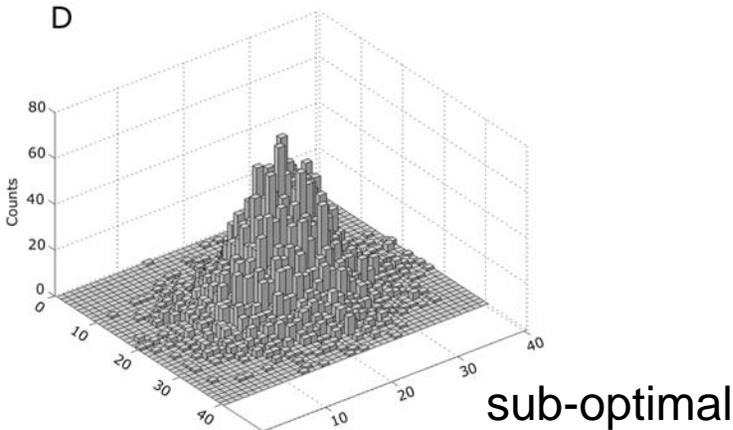
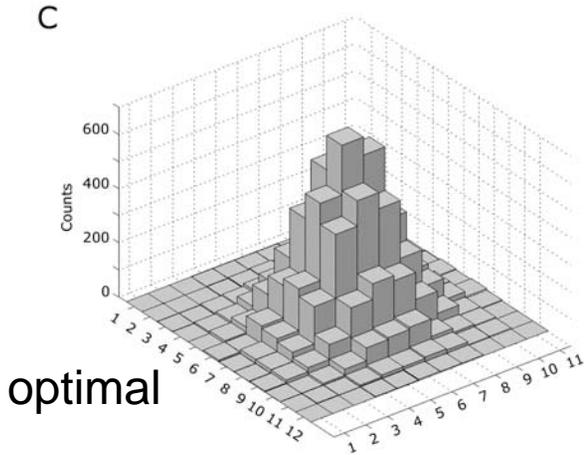
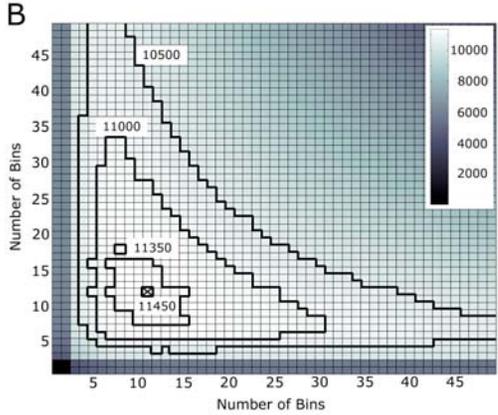
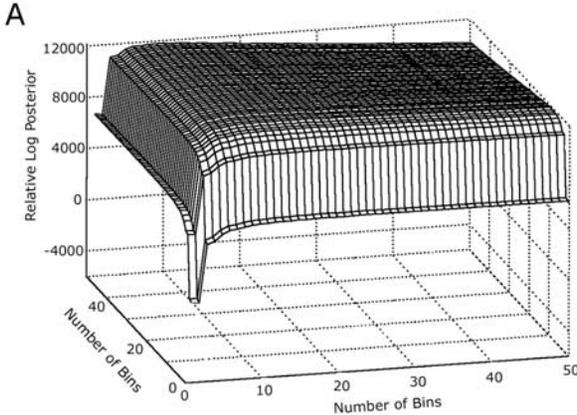
$N = 10000, M = 10$



The histogram should contain only details warranted by the data.

Extendable to Multi-Dimensional Densities

$N = 10000$
 $M = 11, 12$



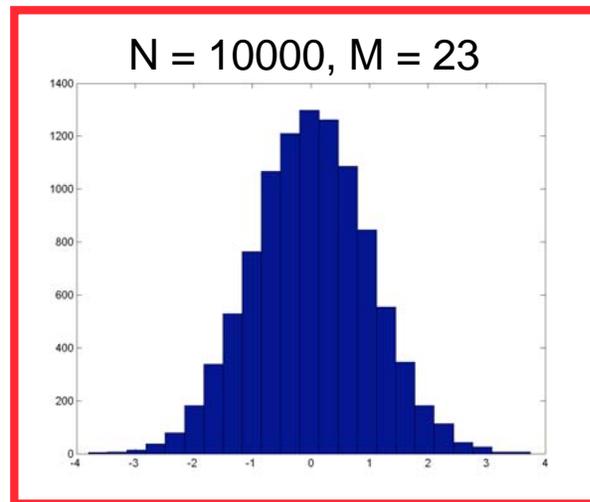
Extendable to multi-dimensional histograms

Entropy Estimation

Entropy estimation is relatively easy with a constant-piecewise model

$$H = - \sum_i p_i \log p_i$$

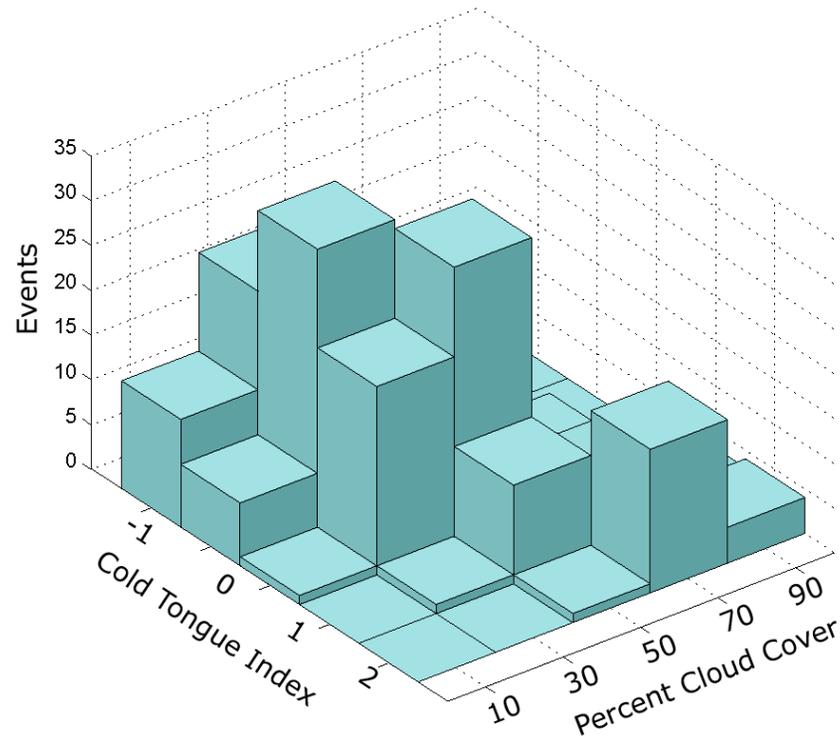
```
H = -sum(p .* (log(p) - log(vol)));
```



Entropy Estimation

And also in higher-dimensions...

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y)$$



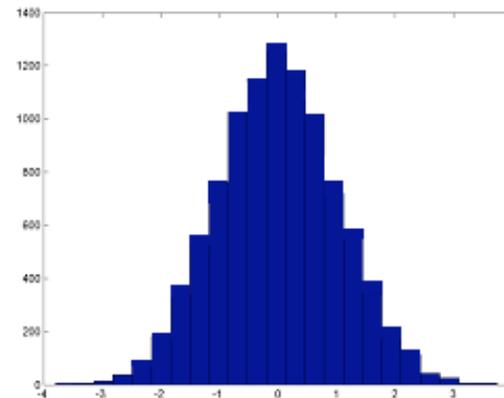
How to Obtain the Uncertainties

To calculate the uncertainties in the entropy estimates, one must first realize that we are uncertain as to the bin probabilities of the probability density model.

By sampling a set of bin probabilities, we obtain a set of probable density functions, along with a set of probable entropies.

$$p(\partial, M | \mathbf{d}, I) \propto \left(\frac{M}{V}\right)^N \frac{\Gamma\left(\frac{M}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^M} \pi_1^{n_1 - \frac{1}{2}} \pi_2^{n_2 - \frac{1}{2}} \dots \pi_{M-1}^{n_{M-1} - \frac{1}{2}} \left(1 - \sum_{k=1}^{M-1} \pi_k\right)^{n_M - \frac{1}{2}}$$

From this set of probable entropies, we can compute the mean and variance. Thus quantifying both the entropy and our uncertainty.



Entropies from Sampling

This shows some of the results from sampling from the posterior probability and computing the entropies.

The data was from a Gaussian distribution with $\mu = 0$, $\sigma = 1$.

The true entropy is $H_{\text{true}} = 1.419$

$N = 10000$, $M = 24$

50000 Samples

$H = 1.4202$

1.4161

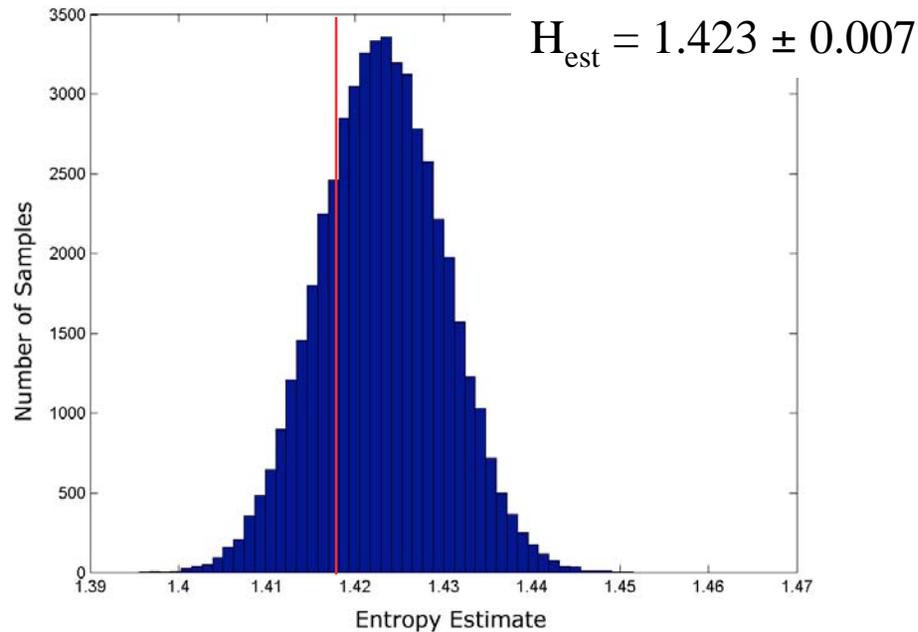
1.4159

...

1.4211

1.4259

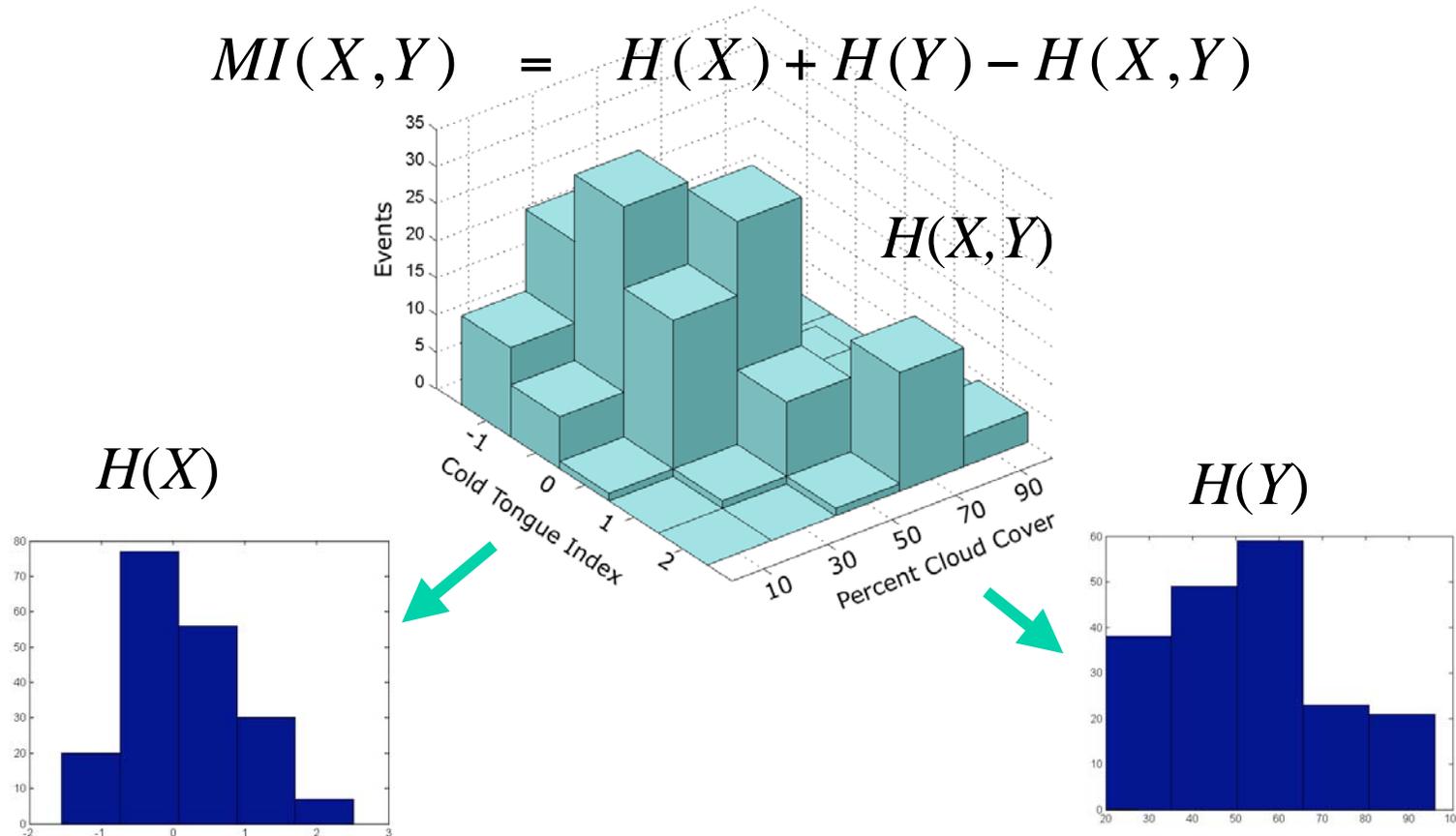
1.4290



Mutual Information

Mutual information requires the estimation of BOTH the two one-dimensional marginal entropies and two-dimensional joint entropy. We can use the same sampling strategy for all cases.

$$MI(X, Y) = H(X) + H(Y) - H(X, Y)$$



Outline

Describing Systems

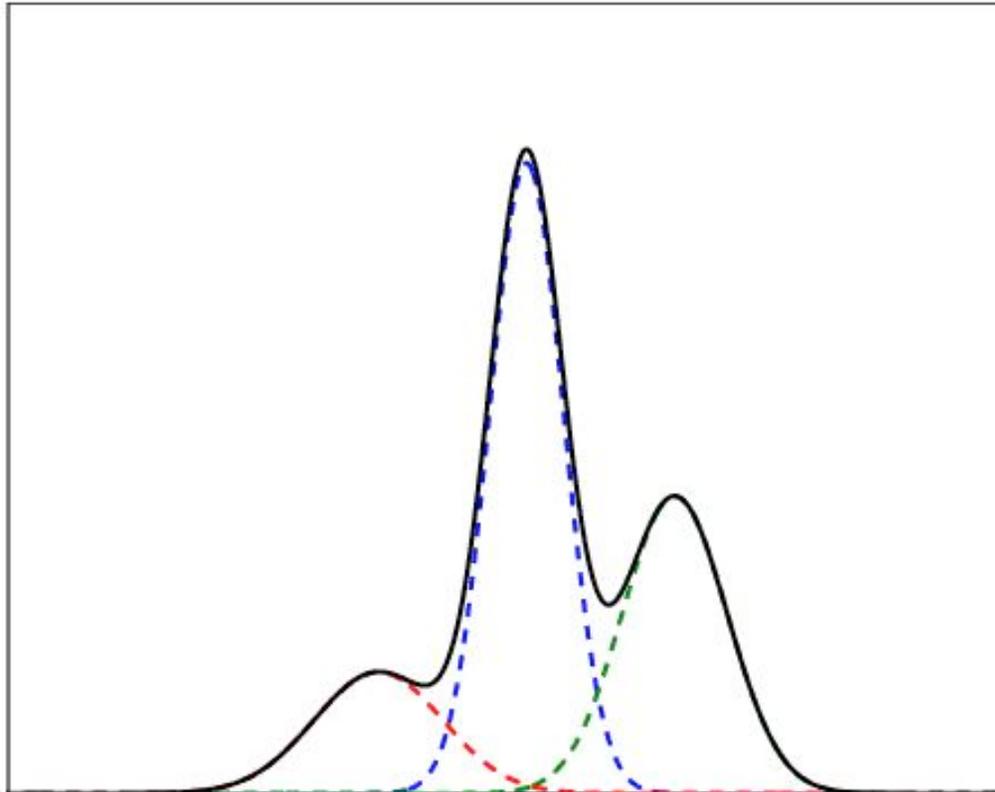
Information Theory

Histogram Methods

Mixture of Gaussians

Deployment Efforts (Ongoing)

Mixture of Gaussians



Another effective density model is the Mixture of Gaussians (MoG)

Figure from R.A. Choudry 2002

Mixture of Gaussians and Nested Sampling

We apply Nested Sampling to this problem to describe the density of the data with a Mixture of Gaussians.

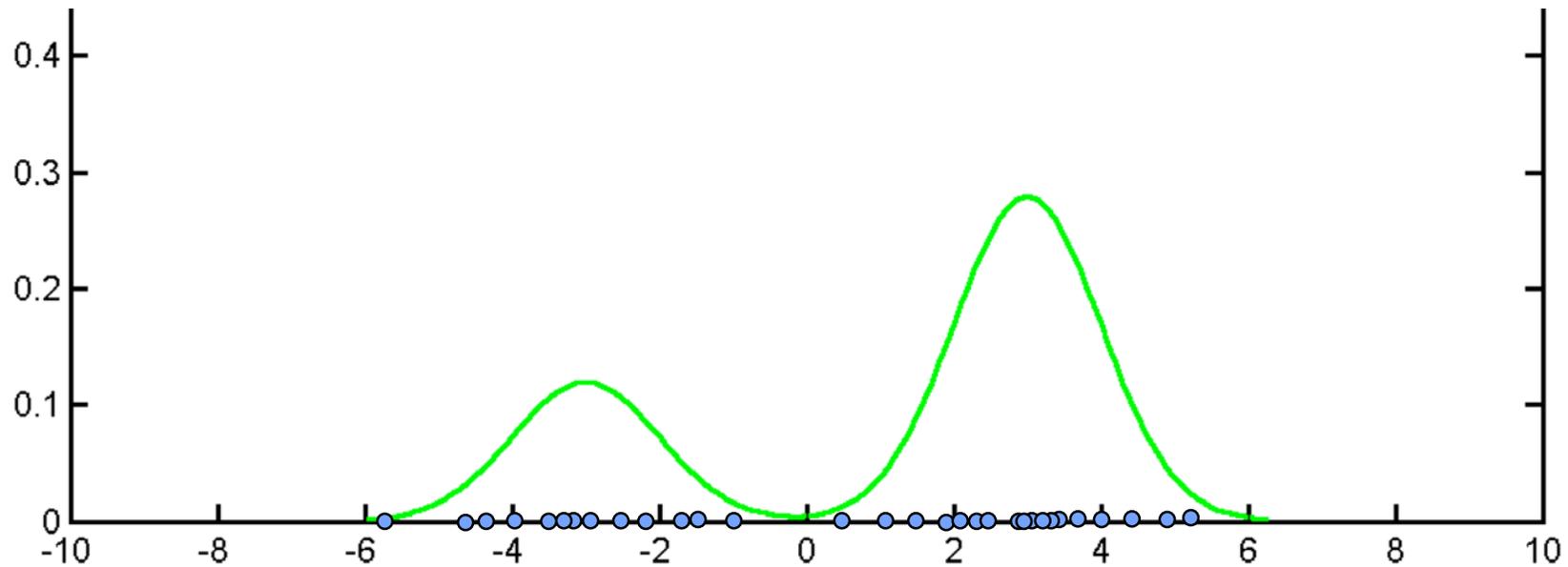
This corresponds to the Likelihood function for a datum x :

$$P(x | M, \mathbf{A}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}) = \frac{1}{Z(\mathbf{A}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}})} \sum_{i=1}^M A_i \text{Exp} \left[-\frac{(x - \mu_i)^2}{2\sigma_i^2} \right]$$

Nested sampling will give us the following:

- 1. A Set of Sample Density Functions**
- 2. A marginal probability for the number of Gaussians M**
- 3. The entropy of the Posterior**

Animation of Nested Sampling in Action



Here the discarded objects (blue) represent a Mixture of Gaussians (MoG) model that converges to the ideal probability density (green) of the samples.

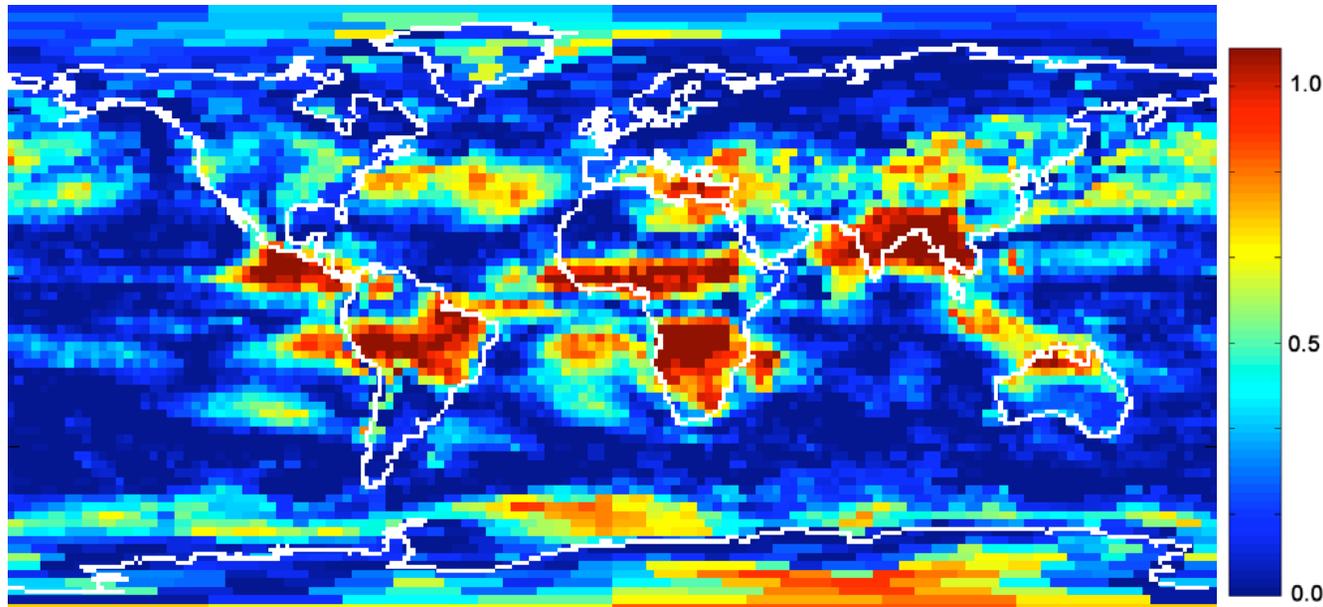
In this figure, the samples are for illustrative purposes and were not the real data sample positions

Examples using Mutual Information

Mutual Information between ISCCP Cloud Cover and Seasonality

Preliminary Mutual Information Results

Mutual Information between ISCCP percent cloud cover and Seasonality.



The data consisted of monthly averages of percent cloud cover resulting in a time-series of 198 months of 6596 equal-area pixels each with side length of 280 km. The analysis was performed pixel-wise so that for each pixel:

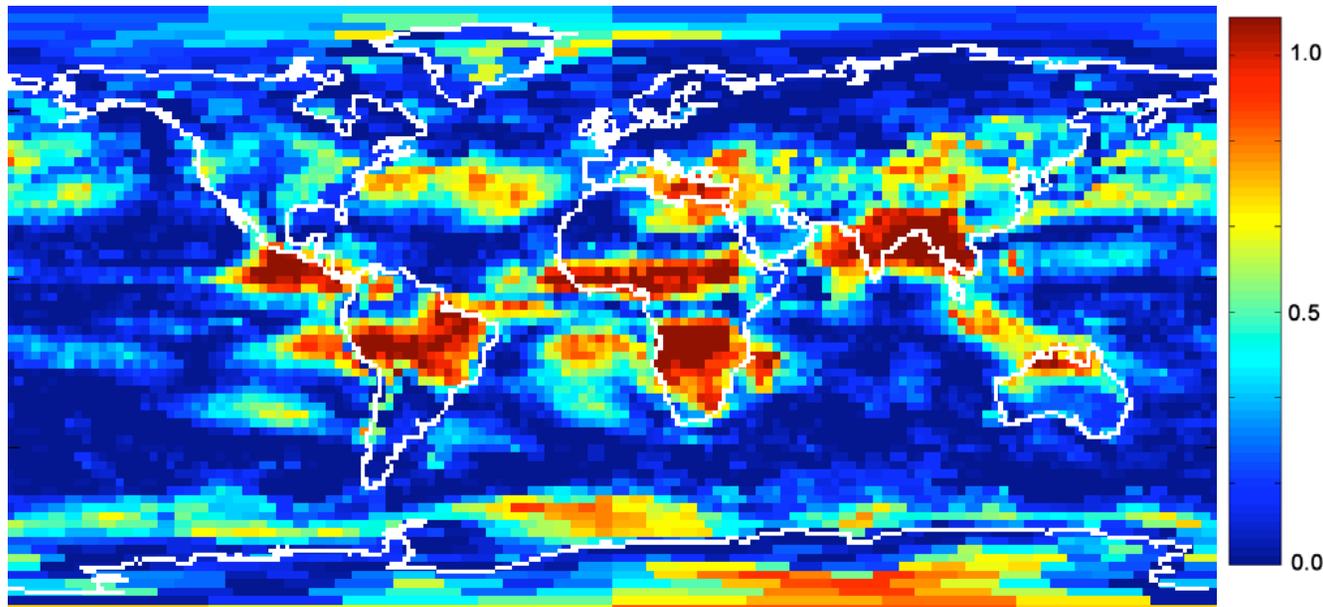
X = cloud cover percentages and Y = month of the year (seasonal state).

The MI was computed for each pixel independently and is color-coded on the map above.

Results based on code by Petr Tichavsky and algorithm by G.A. Darbellay and I. Vajda, IEEE Trans. Info. Theory, 1999

Cloud Cover and Seasonality

Mutual Information between ISCCP percent cloud cover and Seasonality.



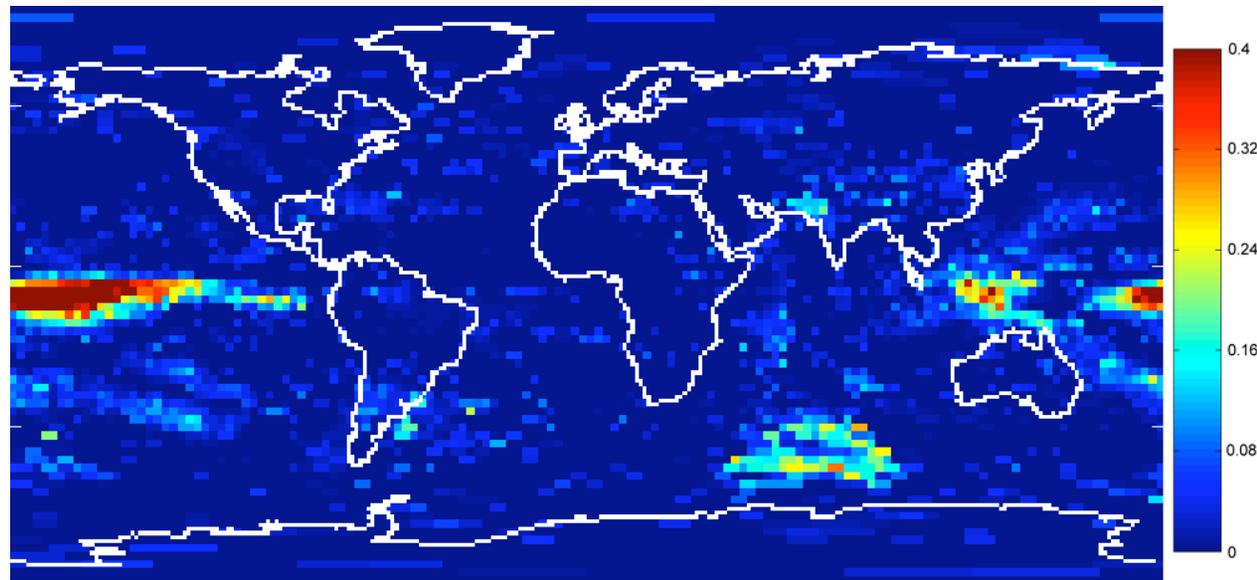
This method finds the Inter-Tropical Convection Zones, The Monsoon Regions, the Sea Ice off Antarctica, and cloud cover in the North Atlantic and Pacific.

This figure can be directly compared to the PCA analysis performed by Rossow et al. 1991, J. Climate, 6:2394-2418.

Mutual Information between ISCCP Cloud Cover and the Cold Tongue Index

Cloud Cover and Cold Tongue Index

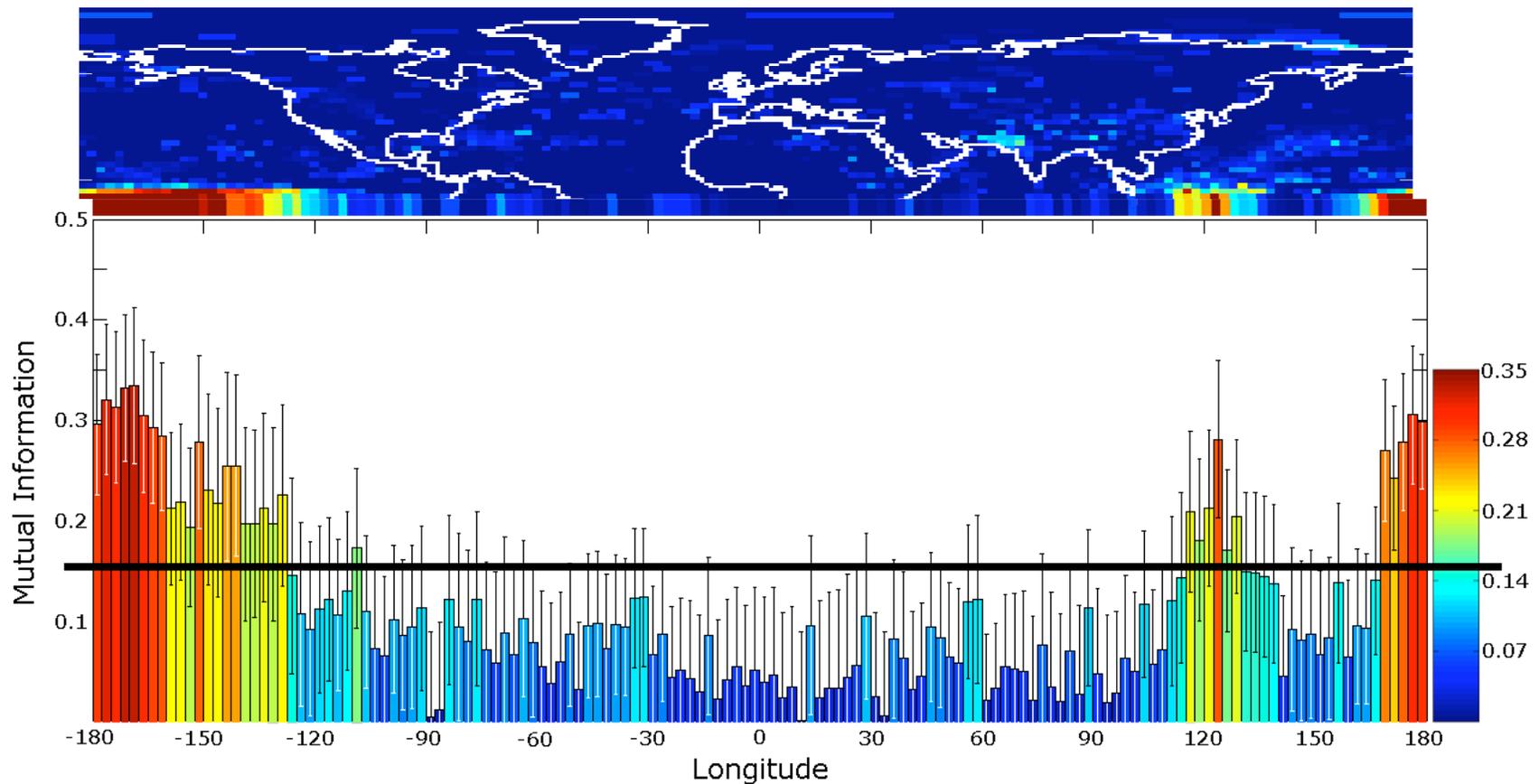
Preliminary mutual Information map between ISCCP percent cloud cover and Cold Tongue Index (CTI), which describes the sea surface temperature anomalies in the eastern equatorial Pacific Ocean (6N-6S, 180-90W) and is indicative of ENSO variability.



Cloud cover data is from ISCCP C2 and CTI data is from T. Mitchell:
http://tao.atmos.washington.edu/pacs/additional_analyses/sstanom6n6s18090w.html

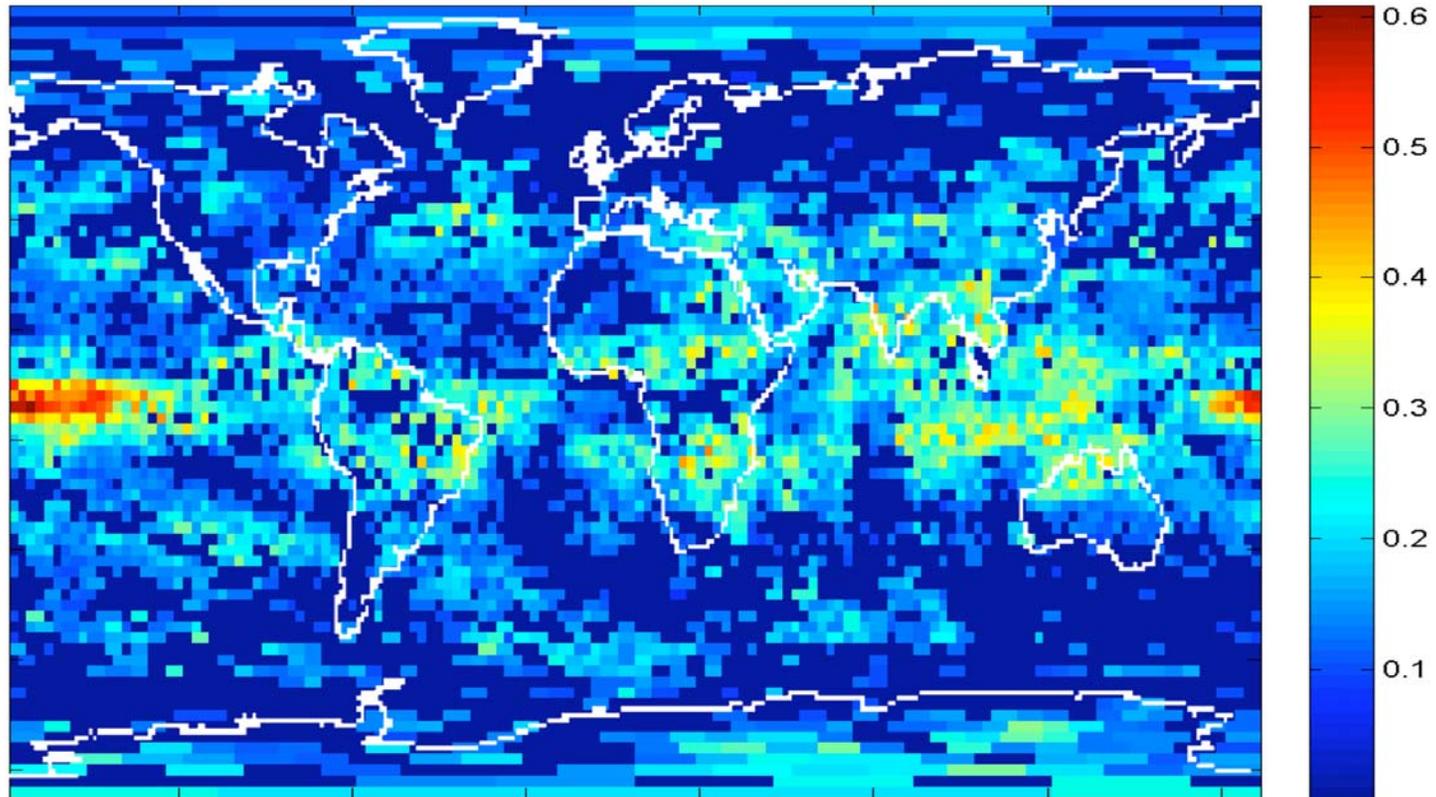
Statistical Significance

Values below line do not indicate statistically significant interactions



ISCCP Cloud Cover and ENSO CTI

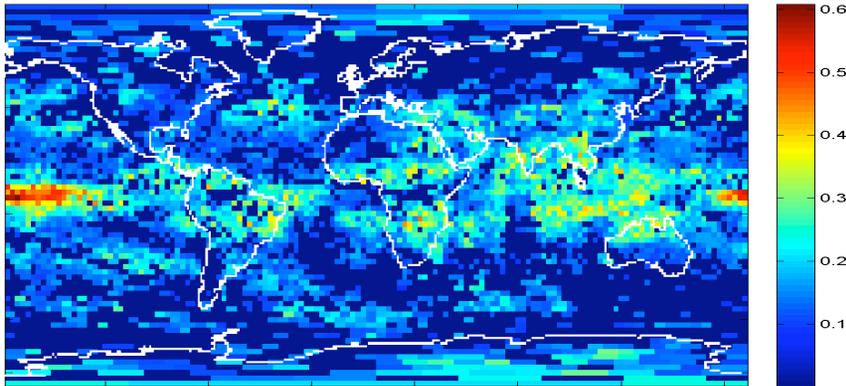
MI MAP OF ENSO SST VS. PERCENT CLOUD COVER
2.0 SIGMA CUTOFF



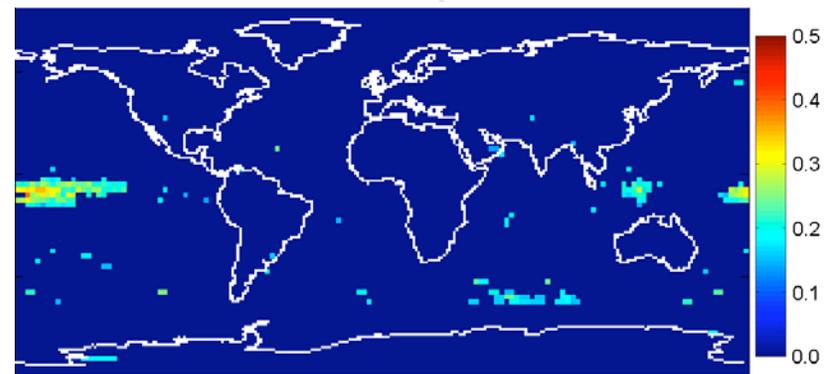
ISCCP Cloud Cover and ENSO CTI

MoG Model

MI MAP OF ENSO SST VS. PERCENT CLOUD COVER
2.0 SIGMA CUTOFF



Histogram Model



The Mixture of Gaussians model reveals a greater impact of ENSO CTI on global cloud cover.

Not only are the equatorial Pacific and Indonesian areas affected, but ENSO activity affects cloud cover in the inter-tropical convection zones in Africa, India and the eastern coast of South America.

Outline

Describing Systems

Information Theory

Histogram Methods

Mixture of Gaussians

Deployment Efforts (Ongoing)

Deployment

We are currently working on the appropriate way to deploy these new computational tools.

The optBINS histogramming package is available online in Matlab at:

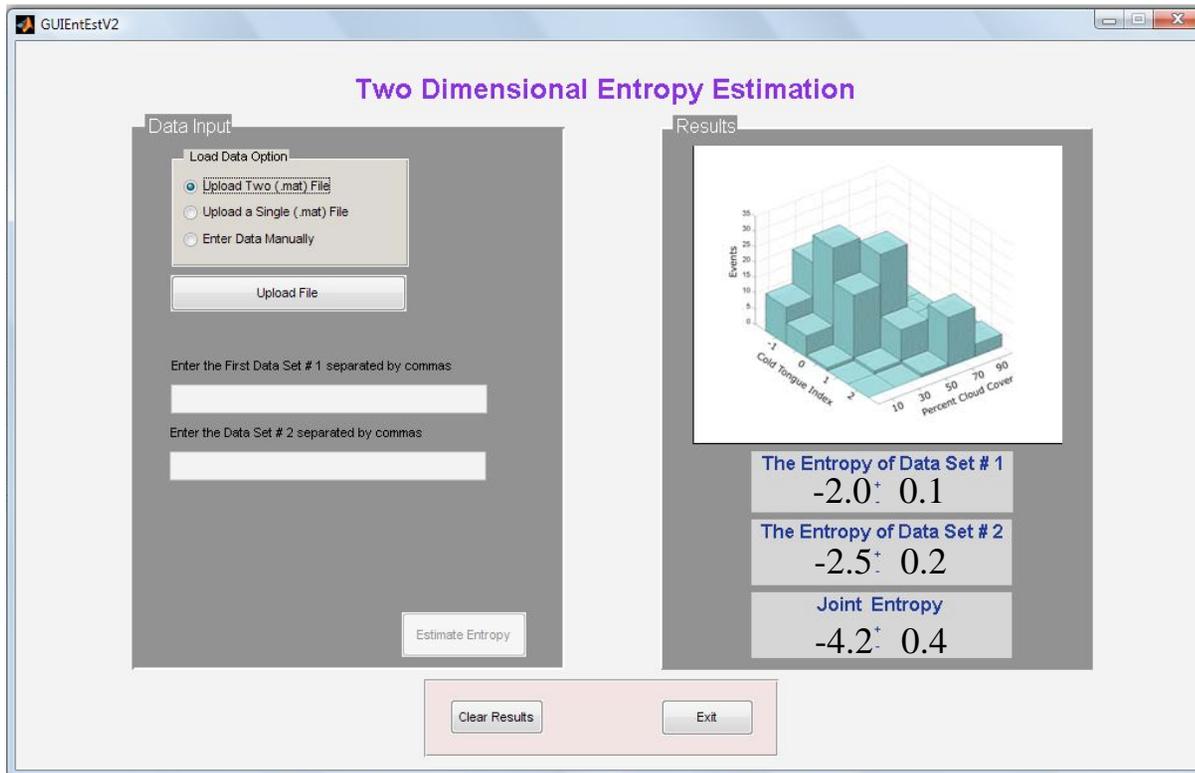
<http://knuthlab.rit.albany.edu/optBINS.html>

We have obtained the Java Builder package from Matlab and are working to design a Graphical User Interface (GUI) Tool that will allow people to use these tools without programming.

We will be uploading these tools to the [NASA AISRP Code and Algorithm Library](#).

Deployment

We are finishing a Web Interface on our servers to provide these entropy estimation tools to those researchers without access to Matlab



<http://knuthlab.rit.albany.edu>

Deployment

We are working with Bill Rossow (CCNY) to work to integrate these tools with his

International Satellite Cloud Climatology Program (ISCCP)

<http://isccp.giss.nasa.gov/>

As well as

Gewex Radiation Panel

<http://grp.giss.nasa.gov/plans.html>

Gewex Cloud System Study–Data Integration

<http://gcss-dime.giss.nasa.gov>

We feel that to do this effectively, we needed to seriously utilize these tools in an experimental Earth science project before making them available in a way that they would be useless.

We are currently working with Bill Rossow on his funded project and to test these tools at TRL 7 before polishing them up for TRL 8.

Broader Impact: Robotics

The conceptual, theoretical, and computational advances made in the course of this work has directly contributed to the application of entropy in experimental design.

The PI has applied this new computational technology to aid in the design of a robotic scientist designed to perform automated sensor placement. Optimal sensor positions are identified by computing the entropy of the predicted results.

Knuth K.H., Erner P.M., Frasso S. 2007.
Designing intelligent instruments. *K.H. Knuth, A. Caticha, J.L. Center, A. Giffin, C.C. Rodriguez* (eds.), *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, Saratoga Springs, NY, USA, 2007, AIP Conference Proceedings 954, American Institute of Physics, Melville NY, In press.

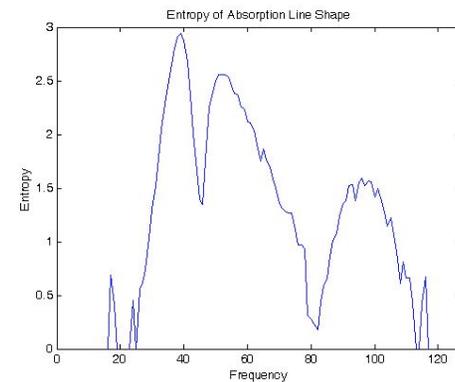
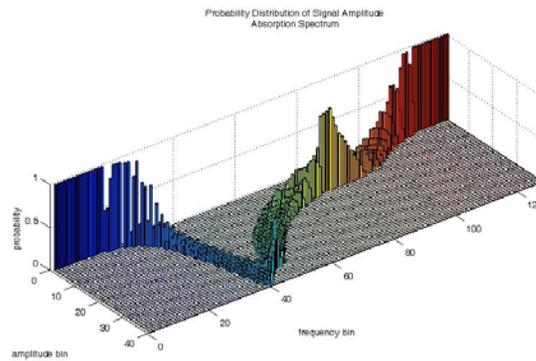
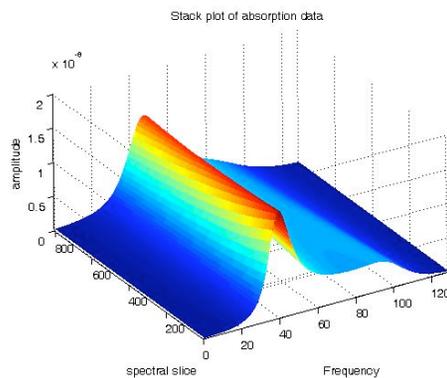


Broader Impact: Spectrometry/Experimental Design

A colleague of the PI, Dr. Keith Earle (University at Albany) is now using entropy-based methods to design Electron Paramagnetic Resonance (EPR) experiments.

These results were presented at the ACERT Workshop on Analysis of ESR Data for Motional Dynamics:

"HFHF ESR and Multifrequency Analysis" by Dr. Keith A. Earle



SUMMARY

Computational techniques that allow us to estimate information-theoretic quantities from data provide us with a unique opportunity to identify relationships among data sets in an automated way, and will be extremely useful in large data sets.

However, it is extremely important that error bars representing our uncertainties be accurately computed. This is made more crucial by the fact that entropies are bounded quantities.

We are working to ensure that these algorithms and techniques will find great use both within and outside the Earth Science Community.

Thanks to

NASA ESTO AIST Program
for their support of

*Estimation of Entropy with Error Bars:
Computing Information-Theoretic Measures of Causality,*
K.H. Knuth (PI)